RECEIVED
CENTRAL FAX CENTER

SEP 25 2006

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Shingo UCHIHASHI et al.

Docket No: CQ10196

Appln. No.: 09/981,735

Group Art Unit: 2614

Confirmation No.: 6166

Examiner: Melur RAMAKRISHNAIAH

Filed: October 19, 2001

For: SYSTEMS AND METHODS FOR COMPUTER-ASSISTED MEETING CAPTURE

DECLARATION OF PRIOR INVENTION IN THE U.S. TO OVERCOME CITED PATENT OR PUBLICATION (37 C.F.R. § 1.131)

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Purpose of Declaration

1. This declaration is to establish completion of the invention in this application in the United States, at a date prior to January 12, 2001, the effective date of the prior patent publication:

2002/0130955 A1 (Pelletier)

2. The person making this declaration is one of the joint inventors and is the most knowledgeable of the joint inventors.

Facts and Documentary Evidence

3. To establish the date of completion of the invention of this application, the following attached documents are submitted as evidence:

FUJI XEROX (FXPAL) Invention Proposal FXPAL-IP-00-013 dated May 31, 2000 and June 1, 2000;

Atty. Docket No. CQ10196
PATENT APPLICATION

DECLARATION OF PRIOR INVENTION IN THE U.S. TO OVERCOME CITED PATENT OR PUBLICATION (37 C.F.R. § 1.131)
U.S. Application No. 09/981,735

Fax Cover Sheet from Richard Domingo of Xerox to Steve Roe together with the attached FUJI XEROX (FXPAL) Invention Proposal FXPAL-IP-00-013, time-stamped February 10, 2001 at 13:55.

4. From these documents, it can be seen that the invention in this application was made:

at least by the date of May 31, 2000, which is a date earlier than the effective date of the reference.

Diligence

5. The attached statement also establishes the diligence of the applicants, from the time of their conception, to a time just prior to the date of the reference up to the filing of this application.

Time of Presentation of the Declaration

6. This declaration is submitted after final rejection, but before or on the same date of filing an appeal. This declaration could not have been submitted earlier, before the Final Rejection, because the reference that is being removed (Pelletier) has been first cited by the Examiner in the Final Rejection. This declaration overcomes all pending claim rejections. Thus, a good and sufficient reason exists why the present declaration and other evidence is necessary and was not earlier presented.

Declaration

7. I am a co-inventor of the subject matter claimed in the present patent application.

At the time I made the subject invention, I was and presently continue to be an employee of FUJI XEROX.

Atty. Docket No. CQ10196
PATENT APPLICATION

DECLARATION OF PRIOR INVENTION IN THE U.S. TO
OVERCOME CITED PATENT OR PUBLICATION (37 C.F.R. § 1.131)
U.S. Application No. 09/981,735

On or about May 31, 2000, I described my invention in a FUJI XEROX internal document FUJI XEROX (FXPAL) Invention Proposal FXPAL-IP-00-013 dated May 31, 2000 and June 1, 2000 and entitled "Computer-assisted meeting capturing system."

Immediately thereafter, the aforesaid Invention Proposal FXPAL-IP-00-013 was submitted to FUJI XEROX legal department. Upon the receipt of the aforesaid Invention Proposal FXPAL-IP-00-013 by the FUJI XEROX legal department, my invention was assigned docket number IP-FA-0013.

After the detailed review and approval process at FUJI XEROX, the aforesaid Invention Proposal FXPAL-IP-00-013 describing my invention was sent to the law firm of Oliff & Berridge, PLC, for preparation and filing. I am informed that the materials describing my invention were sent to by Xerox employee Richard Domingo to Steve Roe of the law firm of Oliff & Berridge, PLC, on February 10, 2001, as indicated by the Fax Cover Sheet from Richard Domingo of Xerox to Steve Roe attached hereto and time-stamped by the facsimile machine February 10, 2001 at 13:55.

The law firm of Oliff & Berridge, PLC proceeded to prepare a patent application and filed a provisional patent application serial number 60/268,908 based on the aforesaid Invention Proposal FXPAL-IP-00-013 on February 16, 2001. The subject patent application claims benefit of priority of the aforesaid provisional patent application serial number 60/268,908.

I hereby declare that all statements made herein of my own knowledge are true, and that all statements made on information and belief are believed to be true; and further that these

Atty. Docket No. CQ10196
PATENT APPLICATION

DECLARATION OF PRIOR INVENTION IN THE U.S. TO
OVERCOME CITED PATENT OR PUBLICATION (37 C.F.R. § 1.131)
U.S. Application No. 09/981,735

statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Signature(s)

A. Inventor(s)

Full name of sole or first inventor: John Boreczky

Inventor(s) signature John Borg

Date 9/23/06 Country of Citizenship USA

Residence Address: 516 DUTTON AVE., SAN LEANDRO, CA 94577

RECEIVED
CENTRAL FAX CENTER

SEP 25 2006

IPFA 0013

FXPAL Invention Proposal FXPAL-IP-00-013

To: Xerox Patent Department
3333 Coyote Hill Road
Palo Alto, CA 94304

From: FX Palo Alto Laboratory, Inc.
3400 Hillview Ave, Building 4
Palo Alto, CA 94304

Proposal Submitted By

Shingo Uchihashi
PAL 0049
shingo@pal.xerox.com
(650) 813-6907

John Boreczky
PAL 0027
johnb@pal.xerox.com
(650) 813-7585

Jonathan Foote
PAL 0046
foote@pal.xerox.com
(650) 813-7071

Manager

Lynn Wilcox
PAL 0020
wilcox@pal.xerox.com
(650) 813-7574

Proposal Title

Computer-Assisted Meeting Capturing System

Brief Description

This invention proposal describes a system which helps users capture meeting events as multimedia documents. It provides intuitive interfaces for camera controls, visual representation of room activities and the system status, thus letting users make better decisions on directing room cameras and selection of camera.

Introduction

Capturing meetings and presentations as multimedia documents provides various opportunities. A captured presentation can be broadcast, allowing remote viewers to participate and share the experience with local viewers. A video recorded meeting allows people to review a meeting that they have attended or to catch up on a meeting that they have missed.

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

The issue is that it is difficult to make a reasonable quality video of a meeting or a presentation. The significant events have to be captured properly by putting focus on them in order to make the video interesting to watch.

Usually, multiple cameras are used to capture many aspects of activities in various kind of meetings. Each camera is able to pan, tilt and zoom so that it can be directed to capture an interesting event with high resolution. Multiple video streams from the cameras are fed to a switcher and the most interesting one is selected at each moment during the meeting. The manipulations of the cameras and the switcher require a lot of know-how and therefore it requires a skilled operator to generate a recording of a meeting.

The skill required to capture a meeting causes a number of problems. It is expensive to hire a skilled person, and small organizations may not be able to afford one. The person may not be available all the time, the operator may be out of the office, or there may be more than one presentation at the same time. Also a restricted meeting may not allow any operators to attend. Thus, it is desirable to make it possible to record meetings without dedicated operators.

Videos shot by non-experts are not interesting to watch for most of the time. A lot of people just put a single camera in a meeting room. This method often fails to capture the significant events without someone driving it. Other people use multiple cameras for recording, but novice users cannot handle them and tend to leave a single video stream recorded for a long time making the resulted video boring.

To produce a useful video record of a meeting, two major tasks need to be accomplished, manipulation of cameras and video stream switching. Recording all the streams reduces the burden of switching but makes camera control more difficult. When recording only one stream, abrupt camera transition is acceptable as long as a video stream from the camera is not included in the final record. It will be less forgiving if everything is kept as a record.

Currently, human decisions and skills are required to handle the above two tasks. Some automated solutions may be available if conditions could be restricted, i.e. fixing a type of a meeting, limiting a number of speakers or restricting the area a speaker can move. However, those restrictions are not applicable in general meeting situations.

Instead of jumping into building an automated system with limited capability, our approach is to provide enough assistance to novices so that they can produce video recordings as well as skilled operators. Participants in meetings or presentations are not restricted in any way, and they do not need to wear any uncomfortable equipment. Even though the system is not a fully automated meeting capturing system, we believe this is a good step toward the goal.

Witnessed and Understood by [signature] *Date* 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/01/00

System Description

Overview

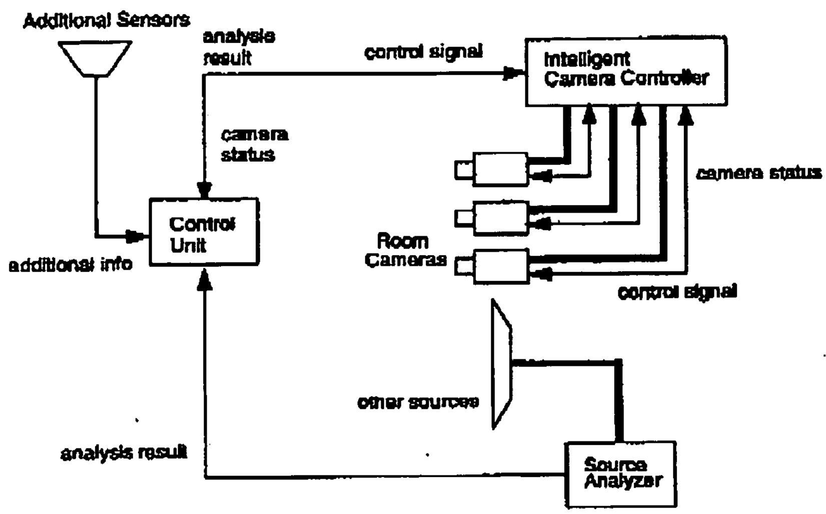

Figure 1 : System Overview for Camera Controlling Components.

Our system consists of three major components as shown in Figures 1 and 2. A control unit allows intuitive camera control and video stream switching using a graphical user interface. The interface displays images from room cameras and other image sources as well as status of the cameras, activities in the conference room and notifications from the system. The intelligent camera controller interprets commands from the user interface and controls the cameras. The controller may receive high level commands for autonomous control of the cameras, such as tracking a person and automatic zoom adjustment. The third component is various analyzers that are scattered over the system. Each analyzer has a small task and sends a result to the control unit to be displayed.

Witnessed and Understood by ______ Date 6/1/00

1. Submitter Signature ______ Date 5/31/00

2. Submitter Signature ______ Date 5/31/00

3. Submitter Signature ______ Date 6/1/00

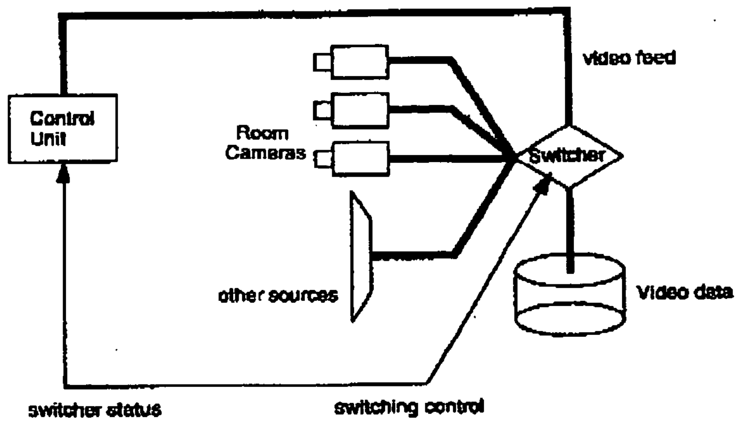

Figure 2 : System Overview for Video Capturing.

Control Unit

Figure 3 is a typical view of a control unit. It has a room view, and a monitor for video streams. Users can control the cameras using either of them using a pen or other pointing device such as a mouse. A room view provides an easy way to directly specify a location where a user wants a camera to be aiming at, while a video stream monitor is suitable for sending incremental instructions to the cameras. The two interfaces complement each other to allow for full camera control.

Witnessed and Understood by Lester D Nelson Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

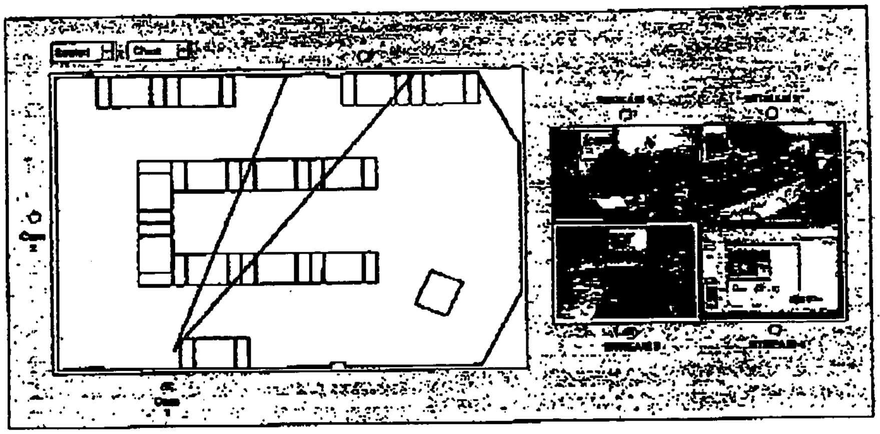

Figure 3 : A typical View of a Control Unit.

Using the room view

The room view shows a floor plan of the conference room as shown in Figure 4. Tapping or clicking on any location of the floor plan will send out a command to the intelligent camera controller to drive a camera to aim at the corresponding location in the actual room. A pull-down menu is used to provide height information that is missing in the floor plan representation. Users can select one of the presets in the menu like "standing" or "seated". Another pull-down menu is provided to set a zoom parameter. The zoom parameter is specified by terms that are widely used among people in broadcast business and also easy to understand by others. Example of the terms are "head", "shoulder", or "chest", each term means to shoot a close-up shot of head, shoulder-up, and chest-up respectively. An advantage of using the terms is that users can relatively specify the zoom parameter not worrying about adjusting it. Other information may be sent to the controller such as

Witnessed and Understood by Lester D. Mehr Date 6/1/00

1. Submitter Signature ______ Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature ______ Date 6/1/00

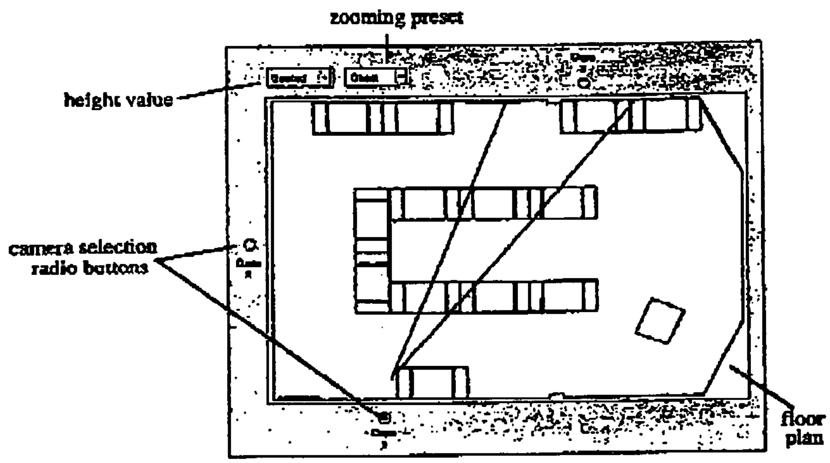

Figure 4 : Room View.

"track a person".

The following is the step-by-step procedure for camera control described above.

1. Select a camera.
2. Select a zoom parameter.
3. Tap or click on the room view.

Various methods are possible for camera selection in step 1 and zoom parameter selection in step 2. Pull-down menu and radio buttons are some of the possible implementation candidates. If a user skips step 1 and step 2, the choices previously selected or default settings will be used. It is also possible to have the system automatically select a camera and zoom setting based on the location selected and the system's knowledge of the room layout.

Another way to manipulate the cameras using the room view is to draw a circle on it. The size and the location of the circle drawn by a gesture is sent to the intelligent camera controller. Then the information is interpreted and generate commands for a camera to shoot the area specified by the circle.

The following is the step-by-step procedure for camera control using the circle gesture.

Witnessed and Understood by ______ Date 6/1/00

1. Submitter Signature ______ Date 5/31/00

2. Submitter Signature ______ Date 5/31/00

3. Submitter Signature ______ Date 6/1/00

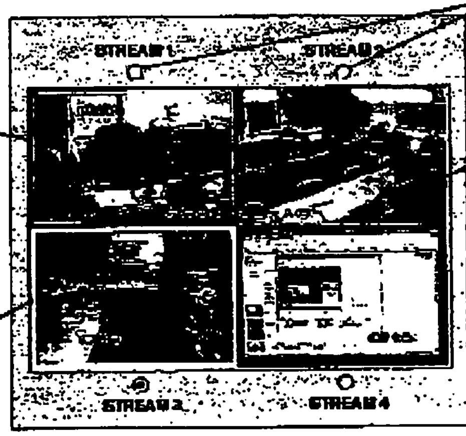

Figure 5 : Stream Monitor.

1. Select a camera.
2. Draw a circle on the room view.

Step 1 may be omitted as discussed before.

Interacting with the stream monitor

The stream monitor is depicted in Figure 5. Users can watch all the video streams sent to a switcher using the stream monitor. The video stream from the camera selected for controlling is hilighted by a white frame. The video stream that is selected by the switcher is hilighted by a colored frame in the monitor. The color of the colored frame changes gradually to indicate a duration that the stream has been selected. The color goes from gray to blue, blue to yellow, and finally to red. Colors with low hue value indicates that the stream is not yet kept long enough and encourage users to hold it more. The red frame strongly suggests users to switch to other streams. The transition of the frame color is shown in Figure 6. Radio buttons located by the stream monitor are used

Witnessed and Understood by Peter J. Mehr Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

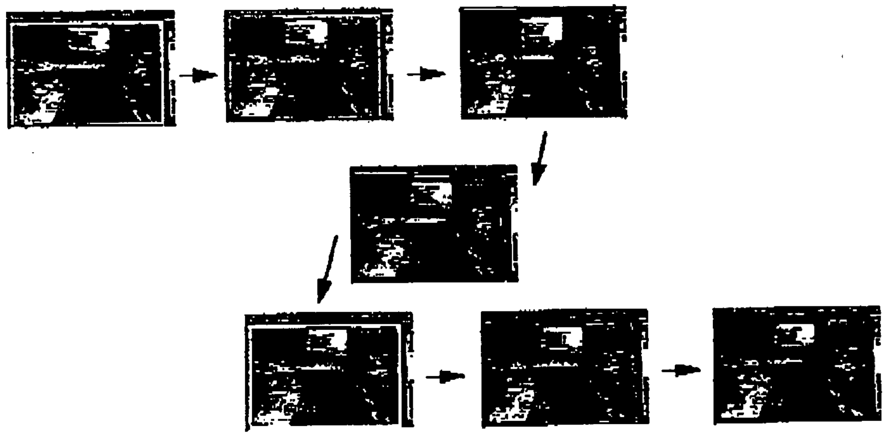

Figure 6 : Frame Color Transition.

to switch among streams. A timer will work as well for experienced producers to show the duration of the same shot instead of colored frames.

The stream monitor can be also used for controlling cameras. Several gestures on an image of a stream allow users to manipulate the associated camera. The commands are drive-up, drive-right, drive-left, drive-down, drive-up-right, drive-up-left, drive-down-right, drive-down-left, zoom-in and zoom-out.

A gesture of a straight line drives a camera to the direction that the line is drawn. The length of the line is used to determine amount of the camera movement. Tapping or clicking on edges of the image can be used to drive the associated camera for a small degree. The direction of the tapped or clicked location from the center of a frame is used.

Other gestures can also be used. For example the zooming gestures from FXPAL IP-99-008 can be used. An inward spiral gesture sends a camera to zoom in, while an outward spiral gesture tells a camera to zoom out. Users can specify a region of interest by making a circle gesture around that area. The gesture is interpreted by the intelligent camera controller and a camera is driven to focus the area.

intelligent camera controller

The role of the intelligent camera controller is to interpret high-level commands from the control unit and to generate low-level commands to drive the cameras. To realize this feature, the control-

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

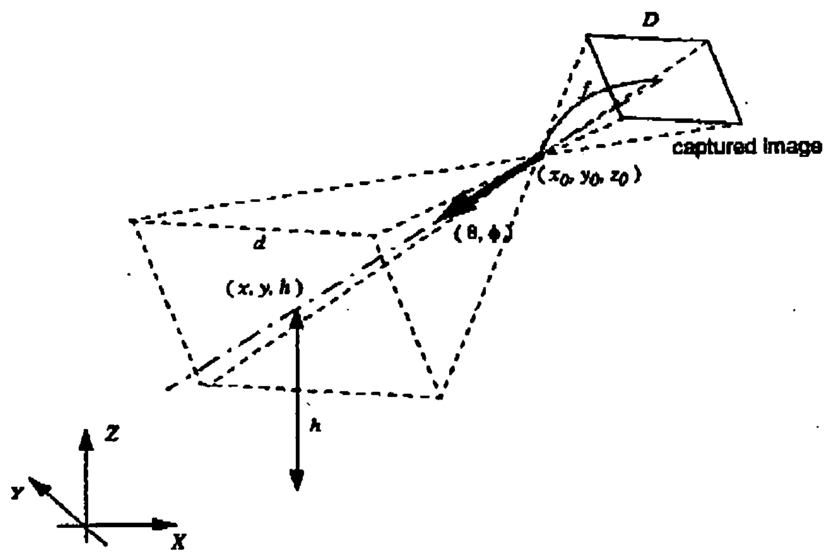

Figure 7 : Camera Coordination.

ler holds geometric information of the conference room as well as parameters to drive the room cameras.

For a pan/tilt capable camera, the center of rotation (x_0, y_0, z_0) can be defined geometrically. Given that parameters to direct the camera to desired angles are known, the camera can be driven any direction (θ, ϕ) to aim at any point in a room within its range of motion. θ is an angle around z-axis, ϕ is an angle from the x-y plane. Also, a zoom capable camera takes a parameter to control a focal length f. By providing an appropriate parameter, the camera can capture pictures of any view angle. Typically, a pan/tilt/zoom capable camera takes three variables v_p, v_t, v_z each to specify amounts of panning, tilting and zooming respectively. The mappings between the variables and the actual camera parameters can be described as in equation (1)(2)(3). If the mappings are linear, equation (1)(2)(3) are rewritten as equation (4), where α_p, α_t, α_f, β_p, β_t, β_f are camera-dependent constant numbers.

$$\theta = \Psi_p(v_p) \tag{1}$$

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

$$\phi = \Psi_t(v_t) \tag{2}$$

$$f = \Psi_f(v_f) \tag{3}$$

$$\begin{pmatrix} \theta \\ \phi \\ f \end{pmatrix} = \begin{pmatrix} \alpha_p & 0 & 0 \\ 0 & \alpha_t & 0 \\ 0 & 0 & \alpha_f \end{pmatrix} \begin{pmatrix} v_p \\ v_t \\ v_f \end{pmatrix} + \begin{pmatrix} \beta_p \\ \beta_t \\ \beta_f \end{pmatrix} \tag{4}$$

A command from the room view of the control unit consists of three information, x-y location, height, and a view angle information. If the command is generated by a tapping gesture, the view angle information is given in an abstract form, such as "head" or "chest". The controller uses an appropriate pre-defined value d to replace the abstract information. For a command by a circle gesture, the size of the circle is used for d. Both a circle and a tapping gesture send one of the preset abstract height value to the controller. This is also replaced by an appropriate pre-defined value h by the controller.

After replacing all the abstract values with real values, the controller has a location (x, y, z) where to aim at and an area that must be covered (d). With the values and camera parameters, the controller determines the variable v_p, v_t, v_z to drive the specified camera to capture images of the specified area. This procedure is done in two steps. First, determine θ, ϕ, and f from the points (x_0, y_0, z_0) and (x, y, h) as equation (5)(6)(7). And then take inverses of equation (1)(2)(3) to get v_p, v_t, v_z.

$$\theta = \operatorname{atan}\frac{y - y_0}{x - x_0} \tag{5}$$

$$\phi = \operatorname{asin}\frac{h - z_0}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} \tag{6}$$

$$f = \frac{D\sqrt{(x - x_0)^2 + (y - y_0)^2 + (h - z_0)^2}}{d} \tag{7}$$

Preset values used to replace abstract values given by the control unit are only good for the first estimation. The intelligent controller adjusts autonomously to meet the original command sent by the control unit. First, the captured images are processed to detect a person. This can be done using various features, such as motion, edges, skin color, or combinations of above. If there is no person detected, the controller stops adjusting the camera position. The camera direction is adjusted to eliminate the gap between the location of the detected person and the ideal location of a person specified by a command.

Once the adjustment is done, the camera captures a person with a desired size. By continuously adjusting the direction of the camera to maintain the person in captured images, the camera can

Witnessed and Understood by [illegible] *Date* 6/1/00

1. Submitter Signature [illegible] Date 5/31/00

2. Submitter Signature [illegible] Date 5/31/00

3. Submitter Signature [illegible] Date 6/1/00

autonomously track the person. This tracking feature can be turned on and off by a command from the control unit.

Intelligent Sensors

A sensor associated with an intelligent component that analyzes signal coming out from the sensor form an intelligent sensor. Various intelligent sensors can be integrated into the system to support users making decision for directing cameras and switching video streams. The outputs from the intelligent sensors are visually present in the control unit so that users can easily recognize them.

A microphone array is one example of sensors. Multiple microphones installed in a conference room can be used to locate where the speaker is. The control unit shows the identified location in the room view by putting a colored blob at the corresponding place. A user can tap on the blob or draw a circle around the blob to drive one of the room cameras to capture the speaker.

Activities in a room can be visually captured using a wide-view-angle camera. IP-99-009 describes one example. A place with the most intensive motions can be easily located by simply taking a difference between every two frames of video from the camera. Such a place is a good candidate of an interesting event. The system detects motion activities in the room, and display the result on the room view of the control unit as colored blobs. Different colors may be used to indicate different degrees of activities. This representation allows users to easily identify where the activities are happening.

Novelty

This invention proposes a system which provides enough assistance for novice users to produce video recordings of meetings and presentations that are interesting to watch. The assistance is implemented as two elements. One element is intuitive user interfaces for controlling room cameras and a switcher to select a video stream to be recorded. The other is automatic notifications of the room activities and system status to the users.

User Interfaces

•Autonomous adjustment of camera positions allows users to select a region of interest with an abstract representation.

•Use of a floor plan for intuitive and direct camera controls.

•The control unit offers an interface for both direct and incremental camera control.

Automatic notifications

•Status of a switcher is notified visually to encourage stream switching at an appropriate frequency.

Witnessed and Understood by Lester D. Mele *Date* 6/1/00

1. Submitter Signature [signature] *Date* 5/31/00

2. Submitter Signature John Boreczky *Date* 5/31/00

3. Submitter Signature [signature] *Date* 6/1/00

•Visual feedback of activities in a conference room is used as a quick overview.

Related invention proposals, patents, publications or commercial work

Other systems have been proposed for automatic recording of presentations. Bianchi presented an fully automated system for televising and/or recording a presentation[2]. Mukhopadhyay, et al. proposed a similar system for automatic lecture capturing system [6]. Their system combines two video streams, a close-up of the lecturer and a slide view, after the lecture. The both systems works only under limited conditions; the application is limited to a presentation given by a single speaker. Instead of building a system with limited use, our approach is to provide enough assistance so that novice users can record any type of presentations or meetings, thus eliminating needs for a dedicated operator.

FXPAL IP-99-008 and [3] describe user interfaces for controlling cameras and selecting views based on having a single panoramic camera view. Providing intuitive interfaces is important for non-experts. Bernier et al. describes an automatic camera zooming system using a face recognition technique. This allows users not to go into details for adjusting cameras. Huang et al. reported an effort to build an automated camera control system[5]. Although their system is still in its early stage and no clear path to a complete system is shown at this moment, their use of a wide-view camera to observe activities in a room may be suitable in this invention too.

Cruz et al. presented a system which broadcasts a lecture using multiple video streams. Users can make a decision to pick the most interesting stream[4]. They solely use personal interest as the criteria for the video stream selection. It is concerned whether the output video is interesting or informative for general viewers.

1 Bernier, O., Collobert, M., Feraud, R., Lemaire, V., Vialet, J. E. and Collobert, D., "MULTRAK: A System for Automatic Multiperson Localization and Tracking in Real-Time," in Proc. ICIP '98, pp. 136-139, 1998.

2 Bianchi, M., "AutoAuditorium: a Fully Automatic, Multi-Camera System to Televise Auditorium Presentation," Joint DARPA/NIST Smart Spaces Technology Workshop, Gaithersburg, MD, July 1998. http://www.AutoAuditorium.com/nist/autoaud.html

3 Chiu, P., Kapuskar, A., Reitmeier, S. and Wilcox, L., "NoteLook: Taking Notes in Meetings with Digital Video and Ink," in Proc. ACM Multimedia '99, pp. 149-158, 1999. ✓

4 Cruz, G. and Hill, R., "Capturing and Playing Multimedia Events with STREAMS," in Proc. ACM Multimedia '94, pp. 193-200, 1994.

5 Huang, Q., Cui, Y. and Samarasekera, S., "Content Based Active Video Data Acquisition Via Automated Cameramen," in Proc. ICIP'98, pp. 808-812, 1998.

6 Mukhopadhyay, S. and Smith, B., "Passive Capture and Structuring of Lectures," in Proc. ACM Multimedia '99, pp. 477-487, 1999.

Witnessed and Understood by [signature] *Date* 6/1/00

1. Submitter Signature [signature] *Date* 5/31/00

2. Submitter Signature [signature] *Date* 5/31/00

3. Submitter Signature [signature] *Date* 6/1/00

Has invention been built, made, run, or tested?

The invention has been partially implemented and tested for experimental use.

Is the invention used in a current product(s) or planned for use in a future product(s)?

Not at this time.

Dates of any previous or planned future disclosures external to Xerox

None

Source of outside funding

None

Witnessed and Understood by Lester D. Nelson Date 6/1/00

1. Submitter Signature ________ Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature ________ Date 6/1/00

1055947

SJR

17

THE DOCUMENT COMPANY
XEROX

Fax Cover Sheet

To: Steve Roe - [redacted]

Recipient's Facsimile Number: 703-836-2787

From: Rich Domingo

Date: 2/10/01

Number of pages (including cover): 14

Special Instructions:

I will have my Admin call you with a docket no. for this case. Thanks Rich

Please call to confirm receipt: ______

Telephone Number: ______

Xerox Telecopier Fax Number: (650) 812-4274

****

Xerox Corporation
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
(650) 812-4000

IPFA 0013

FXPAL Invention Proposal FXPAL-IP-00-013

To: Xerox Patent Department
3333 Coyote Hill Road
Palo Alto, CA 94304

From: FX Palo Alto Laboratory, Inc.
3400 Hillview Ave, Building 4
Palo Alto, CA 94304

Proposal Submitted By

Shingo Uchihashi
PAL 0049
shingo@pal.xerox.com
(650) 813-6907

John Boreczky
PAL 0027
johnb@pal.xerox.com
(650) 813-7586

Jonathan Foote
PAL 0046
foote@pal.xerox.com
(650) 813-7071

Manager

Lynn Wilcox
PAL 0020
wilcox@pal.xerox.com
(650) 813-7574

Proposal Title

Computer-Assisted Meeting Capturing System

Brief Description

This invention proposal describes a system which helps users capture meeting events as multimedia documents. It provides intuitive interfaces for camera controls, visual representation of room activities and the system status, thus letting users make better decisions on directing room cameras and selection of camera.

Introduction

Capturing meetings and presentations as multimedia documents provides various opportunities. A captured presentation can be broadcast, allowing remote viewers to participate and share the experience with local viewers. A video recorded meeting allows people to review a meeting that they have attended or to catch up on a meeting that they have missed.

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

The issue is that it is difficult to make a reasonable quality video of a meeting or a presentation. The significant events have to be captured properly by putting focus on them in order to make the video interesting to watch.

Usually, multiple cameras are used to capture many aspects of activities in various kind of meetings. Each camera is able to pan, tilt and zoom so that it can be directed to capture an interesting event with high resolution. Multiple video streams from the cameras are fed to a switcher and the most interesting one is selected at each moment during the meeting. The manipulations of the cameras and the switcher require a lot of know-how and therefore it requires a skilled operator to generate a recording of a meeting.

The skill required to capture a meeting causes a number of problems. It is expensive to hire a skilled person, and small organizations may not be able to afford one. The person may not be available all the time, the operator may be out of the office, or there may be more than one presentation at the same time. Also a restricted meeting may not allow any operators to attend. Thus, it is desirable to make it possible to record meetings without dedicated operators.

Videos shot by non-experts are not interesting to watch for most of the time. A lot of people just put a single camera in a meeting room. This method often fails to capture the significant events without someone driving it. Other people use multiple cameras for recording, but novice users cannot handle them and tend to leave a single video stream recorded for a long time making the resulted video boring.

To produce a useful video record of a meeting, two major tasks need to be accomplished, manipulation of cameras and video stream switching. Recording all the streams reduces the burden of switching but makes camera control more difficult. When recording only one stream, abrupt camera transition is acceptable as long as a video stream from the camera is not included in the final record. It will be less forgiving if everything is kept as a record.

Currently, human decisions and skills are required to handle the above two tasks. Some automated solutions may be available if conditions could be restricted, i.e. fixing a type of a meeting, limiting a number of speakers or restricting the area a speaker can move. However, those restrictions are not applicable in general meeting situations.

Instead of jumping into building an automated system with limited capability, our approach is to provide enough assistance to novices so that they can produce video recordings as well as skilled operators. Participants in meetings or presentations are not restricted in any way, and they do not need to wear any uncomfortable equipment. Even though the system is not a fully automated meeting capturing system, we believe this is a good step toward the goal.

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/01/00

System Description

Overview

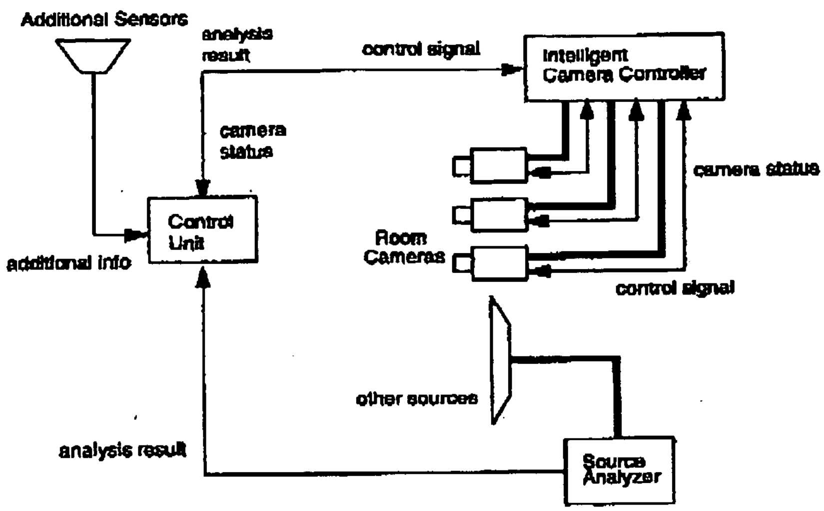

Figure 1 : System Overview for Camera Controlling Components.

Our system consists of three major components as shown in Figures 1 and 2. A control unit allows intuitive camera control and video stream switching using a graphical user interface. The interface displays images from room cameras and other image sources as well as status of the cameras, activities in the conference room and notifications from the system. The intelligent camera controller interprets commands from the user interface and controls the cameras. The controller may receive high level commands for autonomous control of the cameras, such as tracking a person and automatic zoom adjustment. The third component is various analyzers that are scattered over the system. Each analyzer has a small task and sends a result to the control unit to be displayed.

Witnessed and Understood by Lester D. Mel Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

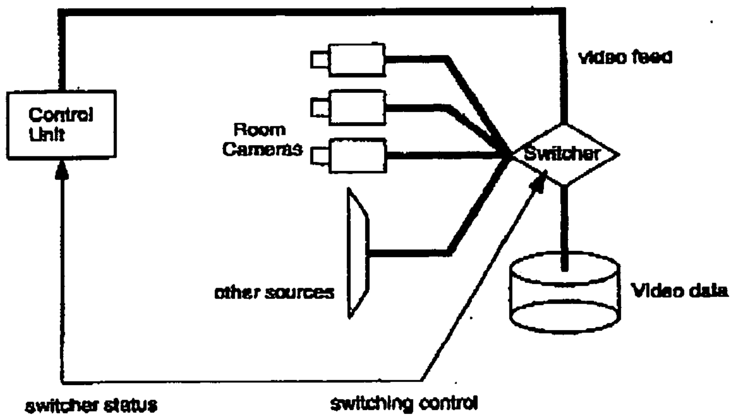

Figure 2 : System Overview for Video Capturing.

Control Unit

Figure 3 is a typical view of a control unit. It has a room view, and a monitor for video streams. Users can control the cameras using either of them using a pen or other pointing device such as a mouse. A room view provides an easy way to directly specify a location where a user wants a camera to be aiming at, while a video stream monitor is suitable for sending incremental instructions to the cameras. The two interfaces complement each other to allow for full camera control.

Witnessed and Understood by Lester D. Nelson Date 6/1/00

1. Submitter Signature ________ Date 5/31/00

2. Submitter Signature ________ Date 5/31/00

3. Submitter Signature ________ Date 6/1/00

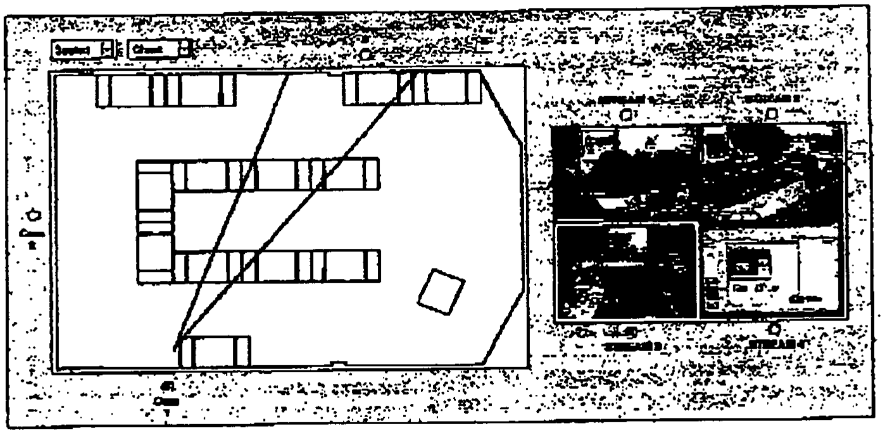

Figure 3 : A typical View of a Control Unit.

Using the room view

The room view shows a floor plan of the conference room as shown in Figure 4. Tapping or clicking on any location of the floor plan will send out a command to the intelligent camera controller to drive a camera to aim at the corresponding location in the actual room. A pull-down menu is used to provide height information that is missing in the floor plan representation. Users can select one of the presets in the menu like "standing" or "seated". Another pull-down menu is provided to set a zoom parameter. The zoom parameter is specified by terms that are widely used among people in broadcast business and also easy to understand by others. Example of the terms are "head", "shoulder", or "chest", each term means to shoot a close-up shot of head, shoulder-up, and chest-up respectively. An advantage of using the terms is that users can relatively specify the zoom parameter not worrying about adjusting it. Other information may be sent to the controller such as

Witnessed and Understood by Lester D. Nelson Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

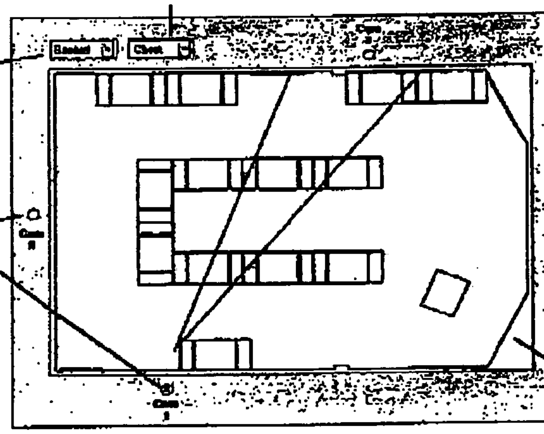

Figure 4 : Room View.

"track a person".

The following is the step-by-step procedure for camera control described above.

1. Select a camera.
2. Select a zoom parameter.
3. Tap or click on the room view.

Various methods are possible for camera selection in step 1 and zoom parameter selection in step 2. Pull-down menu and radio buttons are some of the possible implementation candidates. If a user skips step 1 and step 2, the choices previously selected or default settings will be used. It is also possible to have the system automatically select a camera and zoom setting based on the location selected and the system's knowledge of the room layout.

Another way to manipulate the cameras using the room view is to draw a circle on it. The size and the location of the circle drawn by a gesture is sent to the intelligent camera controller. Then the information is interpreted and generate commands for a camera to shoot the area specified by the circle.

The following is the step-by-step procedure for camera control using the circle gesture.

Witnessed and Understood by ______ Date 6/1/00

1. Submitter Signature ______ Date 5/31/00

2. Submitter Signature ______ Date 5/31/00

3. Submitter Signature ______ Date 6/1/00

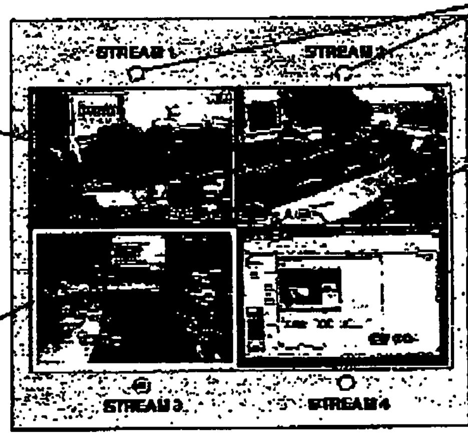

Figure 5 : Stream Monitor.

1. Select a camera.
2. Draw a circle on the room view.

Step 1 may be omitted as discussed before.

Interacting with the stream monitor

The stream monitor is depicted in Figure 5. Users can watch all the video streams sent to a switcher using the stream monitor. The video stream from the camera selected for controlling is hilighted by a white frame. The video stream that is selected by the switcher is hilighted by a colored frame in the monitor. The color of the colored frame changes gradually to indicate a duration that the stream has been selected. The color goes from gray to blue, blue to yellow, and finally to red. Colors with low hue value indicates that the stream is not yet kept long enough and encourage users to hold it more. The red frame strongly suggests users to switch to other streams. The transition of the frame color is shown in Figure 6. Radio buttons located by the stream monitor are used

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

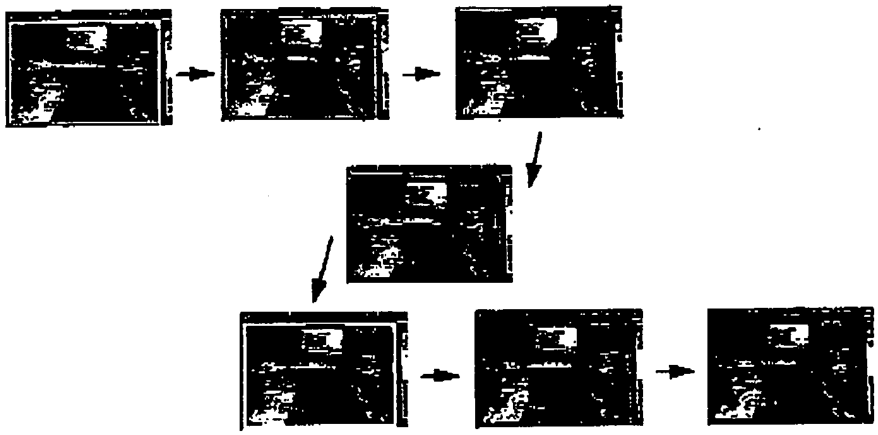

Figure 6 : Frame Color Transition.

to switch among streams. A timer will work as well for experienced producers to show the duration of the same shot instead of colored frames.

The stream monitor can be also used for controlling cameras. Several gestures on an image of a stream allow users to manipulate the associated camera. The commands are drive-up, drive-right, drive-left, drive-down, drive-up-right, drive-up-left, drive-down-right, drive-down-left, zoom-in and zoom-out.

A gesture of a straight line drives a camera to the direction that the line is drawn. The length of the line is used to determine amount of the camera movement. Tapping or clicking on edges of the image can be used to drive the associated camera for a small degree. The direction of the tapped or clicked location from the center of a frame is used.

Other gestures can also be used. For example the zooming gestures from FXPAL IP-99-008 can be used. An inward spiral gesture sends a camera to zoom in, while an outward spiral gesture tells a camera to zoom out. Users can specify a region of interest by making a circle gesture around that area. The gesture is interpreted by the intelligent camera controller and a camera is driven to focus the area.

intelligent camera controller

The role of the intelligent camera controller is to interpret high-level commands from the control unit and to generate low-level commands to drive the cameras. To realize this feature, the control-

Witnessed and Understood by Lester D. Nelson Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

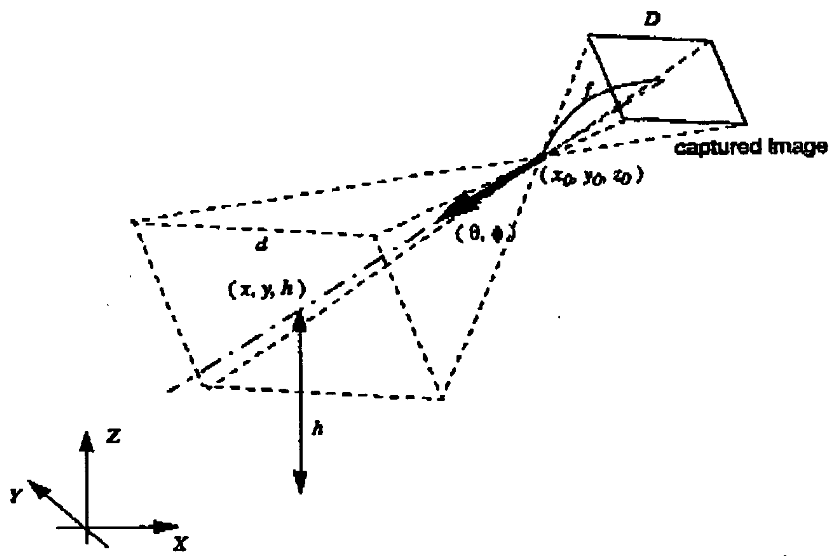

Figure 7 : Camera Coordination.

ler holds geometric information of the conference room as well as parameters to drive the room cameras.

For a pan/tilt capable camera, the center of rotation (x_0, y_0, z_0) can be defined geometrically. Given that parameters to direct the camera to desired angles are known, the camera can be driven any direction (θ, ϕ) to aim at any point in a room within its range of motion. θ is an angle around z-axis, ϕ is an angle from the x-y plane. Also, a zoom capable camera takes a parameter to control a focal length f. By providing an appropriate parameter, the camera can capture pictures of any view angle. Typically, a pan/tilt/zoom capable camera takes three variables v_p, v_t, v_z each to specify amounts of panning, tilting and zooming respectively. The mappings between the variables and the actual camera parameters can be described as in equation (1)(2)(3). If the mappings are linear, equation (1)(2)(3) are rewritten as equation (4), where α_p, α_t, α_f, β_p, β_t, β_f are camera-dependent constant numbers.

$$\theta = \Psi_p(v_p) \quad (1)$$

Witnessed and Understood by Lester D. Mehr *Date* 6/1/00

1. Submitter Signature Steve Uchil... Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature Jonathan Foote Date 6/1/00

$$\phi = \Psi_t(v_t) \quad (2)$$

$$f = \Psi_f(v_f) \quad (3)$$

$$\begin{pmatrix} \theta \\ \phi \\ f \end{pmatrix} = \begin{pmatrix} \alpha_p & 0 & 0 \\ 0 & \alpha_t & 0 \\ 0 & 0 & \alpha_f \end{pmatrix} \begin{pmatrix} v_p \\ v_t \\ v_f \end{pmatrix} + \begin{pmatrix} \beta_p \\ \beta_t \\ \beta_f \end{pmatrix} \quad (4)$$

A command from the room view of the control unit consists of three information, *x-y* location, height, and a view angle information. If the command is generated by a tapping gesture, the view angle information is given in an abstract form, such as "head" or "chest". The controller uses an appropriate pre-defined value d to replace the abstract information. For a command by a circle gesture, the size of the circle is used for d. Both a circle and a tapping gesture send one of the preset abstract height value to the controller. This is also replaced by an appropriate pre-defined value h by the controller.

After replacing all the abstract values with real values, the controller has a location (x, y, z) where to aim at and an area that must be covered (d). With the values and camera parameters, the controller determines the variable v_p, v_r, v_t to drive the specified camera to capture images of the specified area. This procedure is done in two steps. First, determine θ, ϕ, and f from the points (x_0, y_0, z_0) and (x, y, h) as equation (5)(6)(7). And then take inverses of equation (1)(2)(3) to get v_p, v_t, v_f.

$$\theta = \operatorname{atan}\frac{y-y_0}{x-x_0} \quad (5)$$

$$\phi = \operatorname{asin}\frac{h-z_0}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \quad (6)$$

$$f = \frac{D\sqrt{(x-x_0)^2+(y-y_0)^2+(h-z_0)^2}}{d} \quad (7)$$

Preset values used to replace abstract values given by the control unit are only good for the first estimation. The intelligent controller adjusts autonomously to meet the original command sent by the control unit. First, the captured images are processed to detect a person. This can be done using various features, such as motion, edges, skin color, or combinations of above. If there is no person detected, the controller stops adjusting the camera position. The camera direction is adjusted to eliminate the gap between the location of the detected person and the ideal location of a person specified by a command.

Once the adjustment is done, the camera captures a person with a desired size. By continuously adjusting the direction of the camera to maintain the person in captured images, the camera can

Witnessed and Understood by [signature] Date 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature [signature] Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

autonomously track the person. This tracking feature can be turned on and off by a command from the control unit.

Intelligent Sensors

A sensor associated with an intelligent component that analyzes signal coming out from the sensor form an intelligent sensor. Various intelligent sensors can be integrated into the system to support users making decision for directing cameras and switching video streams. The outputs from the intelligent sensors are visually present in the control unit so that users can easily recognize them.

A microphone array is one example of sensors. Multiple microphones installed in a conference room can be used to locate where the speaker is. The control unit shows the identified location in the room view by putting a colored blob at the corresponding place. A user can tap on the blob or draw a circle around the blob to drive one of the room cameras to capture the speaker.

Activities in a room can be visually captured using a wide-view-angle camera. IP-99-009 describes one example. A place with the most intensive motions can be easily located by simply taking a difference between every two frames of video from the camera. Such a place is a good candidate of an interesting event. The system detects motion activities in the room, and display the result on the room view of the control unit as colored blobs. Different colors may be used to indicate different degrees of activities. This representation allows users to easily identify where the activities are happening.

Novelty

This invention proposes a system which provides enough assistance for novice users to produce video recordings of meetings and presentations that are interesting to watch. The assistance is implemented as two elements. One element is intuitive user interfaces for controlling room cameras and a switcher to select a video stream to be recorded. The other is automatic notifications of the room activities and system status to the users.

User Interfaces

•Autonomous adjustment of camera positions allows users to select a region of interest with an abstract representation.

•Use of a floor plan for intuitive and direct camera controls.

•The control unit offers an interface for both direct and incremental camera control.

Automatic notifications

•Status of a switcher is notified visually to encourage stream switching at an appropriate frequency.

Witnessed and Understood by Lester D. Mele *Date* 6/1/00

1. Submitter Signature ______ Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature ______ Date 6/1/00

•Visual feedback of activities in a conference room is used as a quick overview.

Related invention proposals, patents, publications or commercial work

Other systems have been proposed for automatic recording of presentations. Bianchi presented an fully automated system for televising and/or recording a presentation[2]. Mukhopadhyay, et al. proposed a similar system for automatic lecture capturing system [6]. Their system combines two video streams, a close-up of the lecturer and a slide view, after the lecture. The both systems works only under limited conditions; the application is limited to a presentation given by a single speaker. Instead of building a system with limited use, our approach is to provide enough assistance so that novice users can record any type of presentations or meetings, thus eliminating needs for a dedicated operator.

FXPAL IP-99-008 and [3] describe user interfaces for controlling cameras and selecting views based on having a single panoramic camera view. Providing intuitive interfaces is important for non-experts. Bernier et al. describes an automatic camera zooming system using a face recognition technique. This allows users not to go into details for adjusting cameras. Huang et al. reported an effort to build an automated camera control system[5]. Although their system is still in its early stage and no clear path to a complete system is shown at this moment, their use of a wide-view camera to observe activities in a room may be suitable in this invention too.

Cruz et al. presented a system which broadcasts a lecture using multiple video streams. Users can make a decision to pick the most interesting stream[4]. They solely use personal interest as the criteria for the video stream selection. It is concerned whether the output video is interesting or informative for general viewers.

1 Bernier, O., Collobert, M., Feraud, R., Lemaire, V., Vialet, J. E. and Collobert, D., "MULTRAK: A System for Automatic Multiperson Localization and Tracking in Real-Time," in Proc. ICIP '98, pp. 136-139, 1998.

2 Bianchi, M., "AutoAuditorium: a Fully Automatic, Multi-Camera System to Televise Auditorium Presentation," Joint DARPA/NIST Smart Spaces Technology Workshop, Gaithersburg, MD, July 1998. http://www.AutoAuditorium.com/nist/autoaud.html

3 Chiu, P., Kapuskar, A., Reitmeier, S. and Wilcox, L., "NoteLook: Taking Notes in Meetings with Digital Vieo Video and Ink," in Proc. ACM Multimedia '99, pp. 149-158, 1999. ✓

4 Cruz, G. and Hill, R., "Capturing and Playing Multimedia Events with STREAMS," in Proc. ACM Multimedia '94, pp. 193-200, 1994.

5 Huang, Q., Cui, Y. and Samarasekera, S., "Content Based Active Video Data Acquisition Via Automated Cameramen," in Proc. ICIP'98, pp. 808-812, 1998.

6 Mukhopadhyay, S. and Smith, B., "Passive Capture and Structuring of Lectures," in Proc. ACM Multimedia '99, pp. 477-487, 1999.

Witnessed and Understood by Lester D. Mohr *Date* 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00

Has invention been built, made, run, or tested?

The invention has been partially implemented and tested for experimental use.

Is the invention used in a current product(s) or planned for use in a future product(s)?

Not at this time.

Dates of any previous or planned future disclosures external to Xerox

None

Source of outside funding

None

Witnessed and Understood by Lester D. Nelson *Date* 6/1/00

1. Submitter Signature [signature] Date 5/31/00

2. Submitter Signature John Boreczky Date 5/31/00

3. Submitter Signature [signature] Date 6/1/00